A Work Project, presented as part of the requirements for the Award of a Master Degree in Economics from the NOVA – School of Business and Economics.

FACTORS THAT INFLUENCE STUDENTS' SECONDARY ENROLMENT DECISIONS IN PORTUGAL

DIOGO HENRIQUE NOGUEIRA REBELO

A Project carried out in the Master in Economics Program, under the supervision of: Ana Balcão Reis, Luís Catela Nunes and Maria do Carmo Seabra

04-01-2021

Abstract

This work project examines the relation between a measure of academic ability and student's decision regarding track and field of studies to follow, when entering high school. The scores obtained in the 9th grade national exams seem to be the greatest predictors of enrolment decisions. While the Portuguese exam score has greater impact in choosing between a vocational program and an academic one, the Mathematics exam score has greater impact on the probabilities of choosing each of the available academic tracks. Besides academic ability, social economic indicators, predominantly parental education are noteworthy predictors for enrolment decisions in secondary school. This effect however is heterogeneous for different parental education levels in predicting track taking probabilities.

Keywords: Economics, Education, Multinominal Logistic Regression, Social Economic Status, Tracking, Vocational Education,

Acknowledgements

I would like to thank Professors Ana Balcão Reis, Luís Catela Nunes and Maria do Carmo Seabra for their orientation, NOVA SBE for providing the tools and DGEEC for providing the data to carry out this project. I dedicate this Work Project to my family and friends.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

Portuguese students, when entering the 10th grade, are for the first time in their lives presented with the opportunity to choose which area they want to pursue their studies. This decision not only constitutes the first step towards a specialization but it also narrows a lot of future possibilities. With a panoply of courses to choose from, and the implications in terms of future outcomes, in terms of dropout rates, college enrolment rates, career possibilities and so on, the decision of which course to choose will be of extreme importance for the future career of the student. A study conducted by DGEEC (2019) showed that in 2016/2017 approximately 80% of students that followed an academic track in the year after the conclusion of secondary school were enrolled in a superior education institute. This value dropped to just 18% for students from professional tracks. Therefore, identifying the underlying factors that influence student's course decisions is a crucial step to model efficient policies for our educational system.

The Portuguese educational system has suffered significant changes throughout the years in order to tackle the low qualifications of the populations when compared to several others European countries. The Portuguese system employs a specific method of educational tracking when students enter high school. Tracking can be described as separating students into different school classes. Betts (2011) identifies two types of tracking, as this separation can be based on academic ability or curriculum focus. The first consists in grouping students by their achievements in a way where over-achievers would be in a class of their one while less-performing students would be grouped together. A curriculum tracking method meanwhile consists in grouping students by their interest and vocational aptitudes. Although there are arguments that can be made in favour of tracking methods, recent studies have presented evidence tracking generates more inequalities (Hanushek and Wössmann, 2006).

The tracking method used in Portugal is exclusive to a hand-full of countries (e.g. Sweden and Spain). Students are given the opportunity to choose from a given set of specialization tracks.

In secondary schools, students can either choose an academic track (more oriented towards a university degree) or can focus on a more practical vocational educational training (VET).

The VET, in Portugal, has been strengthen especially since 2009 when the secondary education was set as the minimum obligatory schooling level. The intuition for the development of this programs lies in studies that show how these type of curriculums may be particularly helpful in engaging students at the risk of dropping out (Rasinski and Pedlow, 1998) while also increasing student's employment chances (Kulik, 1998; Neuman and Ziederman, 1999). This system allows for a very specific type of training during secondary school and therefore, the decision of which track to choose, will have a great impact on the student's future careers.

While apparently a free choice, there are underlying factors that influence student's decisions. This project proposes to answer two different questions. Based on previous literature it is possible to identify certain variables that influence student's decisions. The first question this project analyses is how the change in those variables affect the probability of choosing between an academic track and a VET. The second question this project analyses is how the change in those variables affect the probability of choosing certain courses within the academic track.

2. Literature Review

The majority of literature, regarding factors influencing career path choices, revolves around the analysis of students´ choices concerning their higher education degree. In Portugal, there isn't much literature concerning these factors at high school level, but due to the highly specialized curriculum of Portuguese secondary education, we are able to extrapolate the arguments made in relation to tertiary education degree choice into our analysis.

2.1 Social Economic Status and Parental Influence

Firstly, it is important to clarify that the decision of which path to follow is not entirely in the students' hands. As research shows, parents have the greatest influence on which careers their children choose (Kniveton, 2004). A study conducted in the United States showed that, parents could even influence what major they children would pursue in college (Simpson, 2003). Although parent's decisions and aspirations for their children is a hard indicator to track, we can use indirect ways to evaluate the parent's expectations, for example by recurring to the analysis of their own education and career path. Parent's education and occupation allows to induce their social economic status (SES), which "is perhaps the most well-researched factor affecting educational plans" (Looker and Lowe, 2001). The parental status provides different sources of spillovers to their son's education. A higher status provides a source of cultural capital that originates from high levels of education and exposure to middle and upper class attitudes and values. On the other hand it also provides a source of social capital that relates to the resources available from the connections to others. Although the context is significantly different, studies in the US show this pattern between parental education and the pursuit of post-secondary education (e.g. Mortenson, 1995; Shepard, 1992). In order to make the bridge to our project, we must consider that the academic track, in the Portuguese system, provides greater chances of continuing the education past the mandatory year's stipulated by the state. Therefore, students that follow this track should have higher SES than the ones from a professional track. Heyns' study (1974) showed that in the US a student whose father's education is 1 standard deviation above average is more prone to enrol in academic programs.

The effects of SES has also been researched in regard to specific field's aspirations. Previous studies conducted proposed high correlation between SES and STEM courses (Coleman, 1987; Davis-Kean, 2005; Eccles et al., 1993; Watkins, 1997). The reasoning behind this effect relates to the theoretical assumptions so far followed. Families with higher incomes and levels of education are more likely to live in neighbourhoods with a great variety of opportunities for their students to learn both inside and outside schools (social spillover), which helps to increase their students achievements (Coleman, 1987). The higher academic ability can be associated with higher percentage of STEM course taking (Madigan, 1997; Schneider et al., 1998).

2.2 Gender

Gender plays a big role in the enrolment decisions in secondary education, although these effects are perceived as very complex. Several studies point out that females from western industrialized countries outperform male students academically (e.g Busch, 1995; Gammie et al., 2003) and are more prone to pursue a university diploma (e.g Lowe et al., 1997; Newfoundland, 1998). Also, male students have a tendency to pursue career paths were traditionally they are more well represented, as opposed to women that choose fields of study in which they were previously under-represented (e.g Ayalon, 2003; Beyer et al., 2004). Male students also are more likely to withdraw before finishing high school (Gilbert et al., 1993; Thiessen and Looker, 1999). While the studies previously mentioned point to a rising educational and occupational aspirations in female students, the gender stereotype still persists to influence women on their career path choices as several studies show (e.g. Pearson and Jenkins, 1997; Thiessen and Looker, 1999; Trusty et al., 2000). Gender stereotypes related to career choices, according to literature, can come from several different sources. Family influence and expectations take once again a big role in the choosing of courses.

The teaching methodology may also have an impact on gender choices. STEM fields, for example, are known to promote emphasis on individual results rather than cooperative learning (Strenta et al., 1994; Seymour and Hewitt, 1997). This sort of environment can be especially discouraging to women, who are more prone to prefer cooperative forms of learning (Barker and Garvin-Doxas, 2004).

2.3 Academic Ability

Academic ability is one of the clearest factors that influence educational plans and attainments. A study conducted in the US finds a correlation between test scores and curriculum assignment of around 0.5 (Jencks et al., 1972). In the Portuguese educational system, the general curriculum followed until the 9th grade, can, in that matter, already give important information to students in which specific field their particular abilities will mostly be relevant.

Academic ability, may also serve as way for students to self-select themselves out of courses perceived as of higher difficulty as Spence (1973) pointed out. He argued that low-ability students have a higher perceived psychological cost when opting for an academic track which will, theoretically, be of higher intellectual demand.

In Portugal, one of the objectives of the VET was to engage students reporting risk of dropping out, therefore is expected that students with lower attainment are more prone to enrol in these courses. A study conducted in Portugal showed that scoring higher in the national exams reduces the odd that a student will enrol in a vocational program (Henriques, 2018).

2.4 Ethnicity and/or Immigrant status

The cultural differences between communities can lead to different attitudes regarding education and even the access to information regarding possible career paths (Looker, 2001). There are also issues relating to intolerance and discrimination. It is important to note that in Portugal most of the immigrant community are members of low-income families from the former African colonies. Another problem arising from the immigrant status, comes from the fact that a recent immigrant from a low-income family may not be aware of the eligibility for certain subsidies designed for students, which can create barriers to the school attainment. Studies conducted regarding these issues in Canada pointed out to these problems (e.g. Hurtado et al., 1997; Maxwell et al., 1996).

2.5 Geographical Location inside the country

Regarding the origin of students within the same country, we can relate this issue with discrepancy between students coming from either rural or urban areas or even students reallocating to other regions to pursue higher education. Students from urban areas have the benefit of having most of the times accessible public transportation for these institutes (Looker, 1993). In Portugal , this issue is a bit more complex since there are several Portuguese districts that do not have a University (e.g. Beja, Bragança, Leiria) although they have other forms of post-secondary education (polytechnic schools). Either way, the difference in post-secondary formation opportunities within the region may have a great impact in the track choice of students. Students may not be able to pay the high cost of reallocation when entering a tertiary degree and for that reason do not enrol in courses more oriented towards academic degrees.

2.6 Other factors

There are several other factors that may influence student's decision, but that may neither be that well documented, or only have higher preponderance within specific groups. From these factors, we may point, for example, family structure. Studies have shown significant differences in post-secondary participation rates between dual or single parent households. Families from single parent households show a significant lower attainment of tertiary education (Furr, 1998; Butlin, 1999; Henriques, 2018). Students attitudes regarding education, although more difficult to measure, also play a factor in the student's individual choices. These factors in students may include perceptions of self-esteem and religious beliefs (Dai, 1996).

These factors are not an extensive list of all the factors that may have an impact in career path choice. This project will only cover Portuguese data, therefore the findings are specific to Portugal. Also, since this project analyses a decision regarding career path that is made earlier than most of the literature presented, the results reflect a different scope of previous literature.

By firstly comparing students from academic tracks and vocational tracks it is expected to see the impact mainly of the SES in preferences and also a correlation of weak academic ability with more vocational tracks. The second part of the analysis focused on the impact of the presented variables in the student´s course decision within the academic track may shed some light in the lack of representation of women in more scientific fields and also how the academic ability may skew lower performing students from the perceived more difficult courses.

3. Institutional Setting

Since 2012, the Portuguese educational system has 12 years of mandatory education. After completing mandatory education, students may apply to tertiary education institutes. Public and Private Institutions coexist in all years of education with most students enrolled in public schools. Commonly students start the school from the year they turn 6 and until the 9th grade they follow a general curriculum. There are several moments of evaluation in national level including mandatory 9th grade national exams in the subjects of Portuguese and Mathematics.

The upper-secondary school, which starts in the 10th grade and ends in the 12th grade, allows students to choose a specialization track from different modalities of school curriculums. The Portuguese system contains a regular academic track, a professional track and some other vocational tracks. In this specialized tracking system the decision of which track to choose is solely in the student´s and parent's hands, with no direct ability-group tracking set in place.

The majority of students (72.7% in our sample)[1] choose the academic track which is designed mostly to students who want to pursue tertiary education. Within the academic track there are four different areas for students to choose from: Sciences and Technology; Socio-Economic Sciences; Languages and Humanities; Visual Arts. All the areas are subjected to a general curriculum composed by the teaching of Portuguese, one foreign language, Philosophy and

[1] According to DGEEC (2020) 60.17% of high school students considered "young" are enrolled in school year 2018/2019 in an academic track.

Physical Education. The specialization tracking comes from the specific formation given according to the area chosen by the student. Each area has a specific mandatory subject which is triennial. The mandatory subject for Socio-Economic and Sciences and Technology students is Mathematics A, while for Languages and Humanities it is History A. For students of Visual Arts the mandatory subject is Drawing A. The student's will have to choose 2 optional subjects between three possibilities given, according to the area of their choosing. These subjects will be biennial. In the 12th grade students have also to choose 2 optional courses that may or may not be directly related to the area previously selected by the student. Schools are given some freedom regarding witch optional subjects to offer to students, normally based on the previous demand of the subject or if the school has the necessary trained faculty member to teach the subjects. This freedom may lead to students being constrained in their subject choices.

The professional track is more focused on a faster insertion of students in the market place. It offers a great variety of programs to students to choose from, although again, limited by the specific school offering. The programs have their respective specific formation, as well subjects common to the entirety of the professional track. In the 12th grade, students of this track are offered work-related training and in order to conclude the course, they must present to a panel of juris a project developed in that year as a proof of professional aptitude. In the period of data analysed, students from this track, who wanted to pursue higher education, would have to take the same exams as the students from the academic track, without any specific school preparation to them. This situation dampen severely the probability for a student from this course to follow tertiary education[2].

[2] A recent new Portuguese law, irrelevant for this dataset, was designed to create, in public higher education institutions, specific vacancies for students from this track in order to mitigate this problem.

4. Data and Methodology

To evaluate the factors which influence the career path choice in upper-secondary education, it has been selected a cohort of students from Portuguese continental public schools that are enrolled for the first time in the 10th grade, between 2011/2012 (first year where mandatory education was extended to the 12th grade) to 2016/2017. The data was extracted using the MISI database, provided by the Portuguese Ministry of Education, and contains an extensive set of information of students enrolled in the Portuguese Educational System, from public schools. This dataset, although it lacks important information regarding students from private schools, is very complete and provides not only information about the students' characteristics but also relevant information about their family, particularly their education. In order to complement the data provided from this dataset, information from the ENEB database, which records students' results on the 9th grade national exams in Portuguese and Mathematics, was also used. The data is composed by 450,394 students from academic and professional tracks.

The first analysis will be conducted by applying the Logistic Regression Model to model whether students follow an academic or a professional program. It will also be analysed if for different parental education levels the impact of the covariates is significantly different allowing to infer if for certain higher parental education levels the independent variables have more or less explanatory power. Thereafter, the focus is within students that chose an academic program and use the same set of independent variables to estimate a Multinomial Logistic Regression to model the choice of the four different set of specializations within the academic track.

4.1 Baseline Specification

Table 1 presents the set of variables used in this project, complemented by a short description. The variables were chosen according to the previous literature presented.

Firstly we define the school year t, which is the year in which the student is enrolled for the first time in the 10th grade. Given that the decision of which course to enrol is made before the beginning of year t, most of the variables have been selected in school year t-1.These variables include our proxy for student's academic ability, which are the results of the mandatory Portuguese and Mathematics national exams in the 9^{th} grade. In order to evaluate the SES of a student one can use the number of years of education of both the mother and the father of the student. These variables were also extracted for the students in the year t-1. A variable representing the average parental education was also added.

The variables that represent the parental number of years of education reflect the parents highest education level completed. For parents who have completed the bachelor's degree the number of years of education is equal to 17 (pre-Bologna accords), the completion of a Master degree represents 19 years of education and a PHD would represent 21 years of education.

Other variable used to evaluate student's SES is SASE which is a subsidy given to lower-income families to cover the costs of education including transportation costs, costs related with school materials and lunch costs. There are three different tiers of subsidy given, which vary according to the financial necessities of the familiar aggregate. The variable used for SASE is a dummy equal to 1 if student receives SASE subsidy and 0 if not. Internet and Computer access also reflect the SES of the student. The lack of access to them constitute material depravation and so it has been also included in the covariates two dummies that reflect if students have or not these tools. Since SASE, computer and internet access can vary over time, it is more important to analyse the variables before the decision period.

Other variables included in the analysis are, as mentioned in previous literature, the gender of the student and the place of birth. The age of the student in year t is also an important measure to take into account since, with some possible exceptions, it reflects the number of previous grade retentions of the student. There are also cases where students, showing great academic aptitudes are invited to skip some grades. These cases will also be reflected in the age variable. In order to eliminate some outliers, we have only selected students with ages between 14 and 18, made at the 31 of December of the year they entered the 10th grade, to be part of our sample (a student without retentions would have 15 years).

In order to eliminate possible omitted variable bias it has been included time and region fixed effects. With the inclusion of time fixed effects it is possible to control for changes in national exams difficulty, government legislation changes regarding SASE subsidies, among others. The region fixed effects allow to account for regional differences in professional and academic track availability and proximity to universities. For the span of 6 years that are included in the investigation we have created 6 dummies representing each one of the school years. Also it is used a dummy for 18 Portuguese districts. This allows to eliminate unobservable variables that evolve over time and region but are constant across entities.

Table 1. *Short definition of the set of explanatory variables included in the models estimated.*

Exogenous Variables	Definition
Educy_mom	Number of years of education concluded by the student's mother in year t-1
Educy_dad	Number of years of education concluded by the student's father in year t-1
Score_pt	Variable ranging from 0 to 100 and reflects the student national exam score in the Portuguese subject in year t-1
Score_mat	Variable ranging from 0 to 100 and reflects the student national exam score in the Mathematic subject in year t-1
Internet	Dummy variable equal to 1 if student has internet access at home and 0 if not in year t-1
Computer	Dummy variable equal to 1 if student has computer access at home and 0 if not in year t-1
SASE	Dummy variable equal to 1 if student receives SASE subsidy and 0 if not in year t-1
Portuguese	Dummy variable equal to 1 if student is from Portuguese origin and 0 if not
Female	Dummy variable equal to 1 if student is female and 0 if male
Age	Variable that represents the age of the student in December of year t
School_year: year1, year2, year3, year4, year5, (year6)	Six dummy variables representing the fixed effects for each school year starting by 2011/2012. The omitted category is year6 and represents the school year in 2016/2017
District: dis1, dis2, dis3, dis4, dis5, dis6, dis7, dis8, dis9, dis10, dis11, dis12, dis13, dis14, dis15, dis16, dis17, (dis18)	Eighteen dummy variables representing the fixed effects for each district. The omitted category is dis18 and represents the district of Viseu
Educy_average	Average of Educy_mom and Educy_dad

5 Decision Factors between Academic and Professional Track

5.1 Descriptive statistics

Appendix 1 presents some summary statistics regarding the two groups of students from the academic and professional track. Our sample comprises 72.66% of students enrolled in the academic track whereby the remaining 27.34% were in a professional track.

Students from academic tracks show a significant higher parental education in relation to students from the vocational track. Other Social Economic variables present the same results. 32.46% of students from the academic track receive some form of SASE support from the state. Meanwhile for Vocational track students this number increases to 51.21%. In relation to internet access at home, only 62.33% of the students from vocational program claims to have it, while in the academic track this number increases to 71.88%. Our results are in line with Spence's (1973) findings. Students from vocational tracks show low academic ability when compared to students from academic track. The average scores for Portuguese and Mathematics exams are significantly lower in the Vocational group, with average scores respectively of 44.03% and 33.31% in the exams. Meanwhile in the Academic track, students have an average score of 58.00% in the Portuguese exam and 54.51% in the Mathematics exam. Students of the Vocational track are on average 0.89 years older than the Academic track counterpart. This reflects a higher retention rate for this group of students.

With 54.66% of females in the academic tracks in relation to only 39.87% of females in vocational programs, our observations seem consistent with previous literature in regards to gender dissimilarities in the pursuit of vocational training (Grebennikov, 2009; Lubinski, 1992).

In order to assess if the mean of the explanatory variables are significantly different among the two groups t-tests were conducted for all the researched variables, with the exception of the fixed effects. All the t-test results pointed to the mean of the variables being significantly different among the two groups with a probability of 100%.

5.2 Model estimation

The first regression models the likelihood of a student to choose between an academic and professional track. In our model the i= {1,2, 3,…,N} denotes our set of decision makers, the students. The model used is the following:

$$\begin{aligned} Academic_{it} &= \beta_1 + \beta_2 female_i + \beta_3 Portuguese_i + \beta_4 SASE_{it-1} + \beta_5 Internet_{it-1} + \\ &\beta_6 Computer_{it-} \ + \beta_7 Educy\_average_{it-1} + \beta_8 Score\_pt_{it-1} + \beta_9 Score\_mat_{it-1} + \\ &\beta_{10} Age_{it} + \delta_1 Year1_{it} + \cdots + \delta_5 Year5_{it} + \theta_1 district1_{it} + \cdots + \theta_{17} district17_{it} + \varepsilon_i \end{aligned} \quad (1)$$

We estimate a Logit Model by maximum likelihood instead of a linear regression model using OLS. The underlying problems of the OLS approach relate to the predicted probabilities falling outside the (0, 1) interval increasing the potential bias and inefficiency of the model due to the heteroskedastic standard errors. With a Logit Model we can correct for these errors given that the logit functional form ensures that the probability of success is strictly between zero and one for all values of the explanatory variables. With this model we will be able to predict an unknown probability P of a student to choose an academic track in detriment of a professional track based on the explanatory variables analysed. With the estimators from the Logit Model we will present the average marginal effect (AME) of the explanatory variables in order to analyse how the probability of choosing an academic track changes with the changes from the explanatory variables. The results can be consulted in *Table 2*.

Aligned with our convictions and as the literature implied, parental education has a positive effect in the pursuit of an academic track. The model used predicts that for each additional year of the student's average parental education, the student is 1.2 p.p. more likely to follow an academic track. Other SES variables, as Internet access and SASE show the same trend in predicting the students track. For students with Internet access, our estimation suggests that on average students are 1.5 p.p. more likely to choose an academic track than students who are Internet deprived. Students beneficiary of SASE are on average 2.5 p.p. less likely to follow an academic track. The student having computer, according to the model, do not impact the probability of choosing a track. The variables that measure SES advocate that students from higher economic background are on average more likely to follow the academic track.

Our model estimations also present important conclusions regarding gender participation rates in both tracks. The coefficient of the female variable implies that, ceteris paribus, female students are 5.9 p.p. more likely to choose an academic track than their male counterparts. Since students who follow an academic track are logically more prone to pursue higher education we can infer that women are more likely to pursue a university diploma as our already mentioned literature states (e.g. Lowe et al., 1997; Newfoundland, 1998).

The proxy used for academic ability, in the form of exam scores, gives clear indicators that higher achievers are more likely to follow an academic track. Surprisingly, it seems that the Portuguese national exam is the one with a higher impact in the track decision. On average, for each additional point in the exam, the probability of the student to follow an academic track increases 0.44 p.p.. In addition, the Mathematics exam, also has an important impact in the track decision. On average for each additional point in the exam in a scale of 0 to 100, the probability of the student to follow an academic track increases 0.31 p.p.. The age variable also has a significant impact on the enrolment decision. For each additional year of age, on average, students are 8.1 p.p. less likely to follow an academic track. Considering that the age variable represents, for a significant number of cases, the number of previous retentions, it reflects the academic ability of the student, therefore, the coefficient from our model, is in accordance with previous literature, in the sense that lower academic achievers are more prone to pursue vocational training.

5.3 Additional regressions and statistics

The previous results assumed that the calculated coefficients were equal across the sample. However, that is a strong assumption. In order to test if the coefficients estimated are indeed equal across the sample, two groups have been defined. Group A is constituted by students whose parents average education years is above 12 and Group B is constituted by students whose parents average education years is equal or lower than 12. Using the same set of independent variables model 1A and model 1B (for students from group A and B) were computed. By employing a Chow test (Chow, 1960), *Appendix 2*, it was possible to conclude that the coefficients among the groups were statistically different, therefore the pooled model 1 conclusions are put in check. The results of the regression 1A and 1B can be consulted below.

Table 2. *Results for model 1, 1A and 1B AME predicting participation in Academic track.*

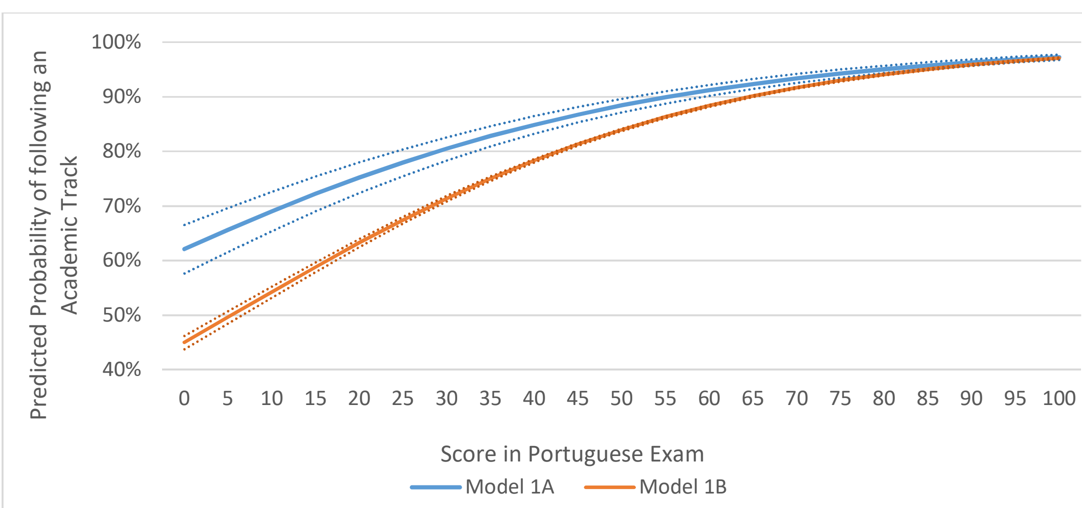
Variables	(1)	(1A)	(1B)
Female	0.0587*** (0.0012)	0.0136*** (0.0017)	0.0716*** (0.0015)
Portuguese	-0.0509*** (0.0030)	-0.0163*** (0.0041)	-0.0613*** (0.0036)
SASE	-0.0246*** (0.0013)	-0.0126*** (0.0025)	-0.0280*** (0.0015)
Internet	0.0150*** (0.0020)	0.0132*** (0.0044)	0.0157*** (0.0023)
Computer	-0.0034 (0.0023)	-0.0012 (0.0047)	-0.0049* (0.0027)
Educy_average	0.0121*** (0.0002)	0.0010** (0.0005)	0.0150*** (0.0003)
Score_pt	0.0044*** (0.0001)	0.0015*** (0.0001)	0.0052*** (0.0001)
Score_mat	0.0031*** (0.0000)	0.0014*** (0.0000)	0.0036*** (0.0000)
Age	-0.0807*** (0.0008)	-0.0375*** (0.0014)	-0.0930*** (0.0010)
Pseudo-R^2	0.2519	0.1825	0.2229
Number of obs.	321,813	71,203	250,610

Standard errors in parentheses
*** p<0.01, ** p<0.05,*p<0.1

Our estimations of model 1A and 1B suggest that students with lower parental education (group B) are more sensitive to changes in the explanatory variables, given the higher absolute value of the coefficients. For students from group A, ceteris paribus, being female increases the probability of following an academic track by 1.4 p.p.. In group B however, ceteris paribus, being female has significant more impact in the probability of following an academic track (7.2 p.p.). The results imply that for lower parental education level being female has more impact in the probability of choosing the course. The results also suggest that for lower levels of parental education a unit increase in average parental years of education will have significant more impact than for students whose parents have on average more than 12 years of education.

In model 1A, the score exam coefficients are also significantly below the coefficients of model 1B. This suggests that for lower parental academic qualifications the score in exams is more preponderant in the course decision. In order to measure the impact of exam scores in the probability of choosing a certain course a different method has been resorted. Using the means of the sample used in model 1 it has been computed the predicted probabilities of choosing the academic track for model 1A and 1B. This allows to understand, for the different models, how the probability of choosing a certain course would change, for the same average individual, for different scores in the national exams (ceteris paribus context). *Figure 1* represents the predicted probabilities for different Portuguese exam scores using model 1A and 1B.

Figure 1. *Predicted probabilities at the mean of choosing the academic track for model 1A and 1B given different Portuguese exam scores (ceteris paribus)*



As expected, given the previous coefficients, for higher Portuguese exam scores the predicted probability of following the academic track increases substantially. For both models, the average individual who's Portuguese score was 100 points would have more than 95% of probabilities of following an academic track. However, it is possible to conclude, that for the same individual, model 1A predicts across the entire Portuguese exam spectrum, higher probability of choosing an academic course than in model 1B. Also, the probability of choosing the academic track varies more in model 1B for different results in the Portuguese exam. This implies that students from higher academic parental background are intrinsically more prone to follow an academic track and are less influenced by exam scores in the choosing of the track. Conducting a similar analysis, but for the mathematics exam, results in similar conclusions that can be consulted in *Appendix 3.*

6. Decision Factors within Academic Track

The following step to our investigation is to use the same explanatory variables and analyse if they have impact in the student's area selection within the academic track, meaning the student, choosing between the four different areas of the Academic track: Sciences and Technology (ST); Socio-Economic Sciences (ES); Languages and Humanities (LH); Visual Arts (VA).

6.1 Descriptive statistics

Appendix 4 presents some summary statistics regarding the composition of the students from each area. The first conclusions that we can take from this preliminary analysis is that the groups are not evenly distributed according to the 4 areas. There is a major predominance of students from the ST area, which corresponds to 55.8% of the students who follow an academic track. The portion of students from the LH is also very significant with 26.5%, with students from ES and VA having respectively 10.5% and 7.2%. The high affluence to the science area is as expected, given that although it is perceived as the area with higher difficulty it also provides the higher panoply of opportunities in terms of university access.

In relation to the explanatory variables the difference among groups is also significant. The area where students have lower parental education is clearly the LH with average of the father education being 8.6 years and the mother education 9.4. The other areas are relatively similar but we can conclude that students from ES have the higher average level of parental education with the average value of mother education being 11.3 and father education being 10.4. The results of our t-tests confirm that all the parental education averages are significantly different among the areas. Other Socio-Economic indicators show the same trend, with for example the percentage of students who receive some form of support through SASE being significantly higher in LH areas where the parental education is lower, while in the ES and ST the percentage of student with this form of support is significantly lower.

The exam scores variables present interesting statistics as well. The Mathematics exam particularly, show an astonishing discrepancy of averages among areas. In this exam the average for students from the ST area is 63.2 points while the average in the LH area is just 36.6 points. For these two areas, while the difference in average is statistically significant (according to the t-test), the difference in the Portuguese exams isn't as much preponderant. In LH the average is 53.7 and the average for the students of ST is 60.7 points. In VA and ES the averages for the exams of Portuguese and Mathematics is respectively 53.6 and 45.3 for VA and 57.1, 58.3 for students from ES. As we can see, students from LH, even in the Portuguese Exam, that should be more related to their vocational aptitudes present lower scores in average than the other groups. This preliminary results suggest that one of the main factors for the area decision within the academic track is the intrinsic academic ability.

The gender composition of the areas indicates that there is still some gender stereotypes that influence the student's decision in which area to enrol. While the academic track has 55% female students, there still seems to exist some resistance to the entry to more predominantly male fields. However the 49.8% of females in the ST field corroborates the findings of Ayalon (2003) and Beyer et al (2004) that pointed to females choosing fields in which they were previously under-represented. Their arguments in relation to male students preferring areas in which they were already well represented also seems correct in the sense that in the VA and LA, the male representation is only 34.9% and 34.8% respectively. However, to look more into their findings we would need to verify the time trend composition by gender of these areas.

6.2 Model estimation

For this second analysis we used a Multinomial Logistic regression Model to predict the likelihood of a student to choose between the four different areas within the academic track. A Multinomial Logit was chosen over a Conditional logit model since regressors do not vary across alternatives (alternative-invariant). In equation (2) below, the i represents the students, and j= {1,2,3,4} represents the mutually exclusive possible choices: the four different academic areas. Under the common Logit Parameterization:

$$\Pr(Area_i = j) = \frac{\exp(X_i\beta_j)}{\sum_{j=1}^{4}\exp(X_i\beta_j)} \qquad (2)$$

where X_i represents the vector of the following explanatory variables: Female, Portuguese, SASE, Internet, Computer, Educy_average, Score_Pt, Score_Mat, Age, Year1…Year5, District 1…District17.

As in the previous models the objective is to investigate by how much each of these variables affect the probability of a student to choose a certain area. The estimations can be consulted in *Table 3* and reflect the AME of the explanatory variables.

Table 3. *Results for model 2: AME predicting participation in the different areas of the academic track.*

	Model 2			
	ST	ES	VA	LH
Female	-0.0994***	-0.0210***	0.0317***	0.0887***
	(0.0018)	(0.0012)	(0.0010)	(0.0015)
Portuguese	-0.0037	0.0016	-0.0014	0.0034
	(0 .0044)	(0.0030)	(0.0022)	(0.0036)
SASE	-0.0085***	-0.0096***	0.0043***	0.0138***
	(0.0020)	(0.0014)	(0.0011)	(0.0017)
Internet	-0.0059*	0.0016	0.0032*	0.0011
	(0.0034)	(0.0024)	(0.0019)	(0.0028)
Computer	0.0118***	-0.0081***	-0.0050**	0.0013
	(0.0038)	(0.0027)	(0.0021)	(0.0031)
Educy_average	-0.0010***	0.0018***	0.0020***	-0.0028***
	(0.0003)	(0.0002)	(0.0001)	(0.0002)
Score_pt	-0.0003***	-0.0014***	-0.0006***	0.0022***
	(0.0001)	(0.0000)	(0.0000)	(0.0001)
Score_mat	0.0081***	0.0009***	-0.0007***	-0.0083***
	(0.0003)	(0.0000)	(0.0000)	(0.0000)
Age	-0.0733***	0.0122***	0.0267***	0.0344***
	(0.0020)	(0.0013)	(0.0008)	(0.0015)
Pseudo-R^2	0.1587			
Number of obs.	256,383			

Standard errors in parentheses
*** p<0.01, ** p<0.05,*p<0.1

Our estimations indicate that gender has a great impact in the probability of following certain areas. Consistent with Pearson and Jenkins (1997) , Thiessen and Looker (1999) and Trusty et al. (2000) one important conclusion that we can take from our data is that gender stereotype still persists and the results suggest that it deeply influences the probability of following a certain area. According to the estimated coefficient of the variable female for the area of ST, on average, females are 9.9 p.p. less likely to choose this area in relation to the others (ceteris paribus). In ES the coefficient is also negative, with female students 2.1 p.p. less likely to choose the area. Meanwhile, females are respectively 3.2 p.p. and 8.9 p.p. more likely to decide for an area in VA and LH than the others. Additional research could be done in this subject by analysing if these coefficients have been changing in past years and if the values are converging to zero. This hypothetical scenario could be argued as proof of the mitigation of gender-based roles and tasks.

Regarding the parental background that translates into SES, our variables of interest, surprisingly, don't seem to have as much impact as in the previous analysis. An additional year of average parental education translates into students to be 0.18 p.p. and 0.20 p.p. more likely to follow the ES and VA areas than the other areas. In reverse with an additional year of average parental education our coefficients predict a decrease of 0.10 p.p. and 0.28 p.p. in the probability of following a ST and a LH area respectively.

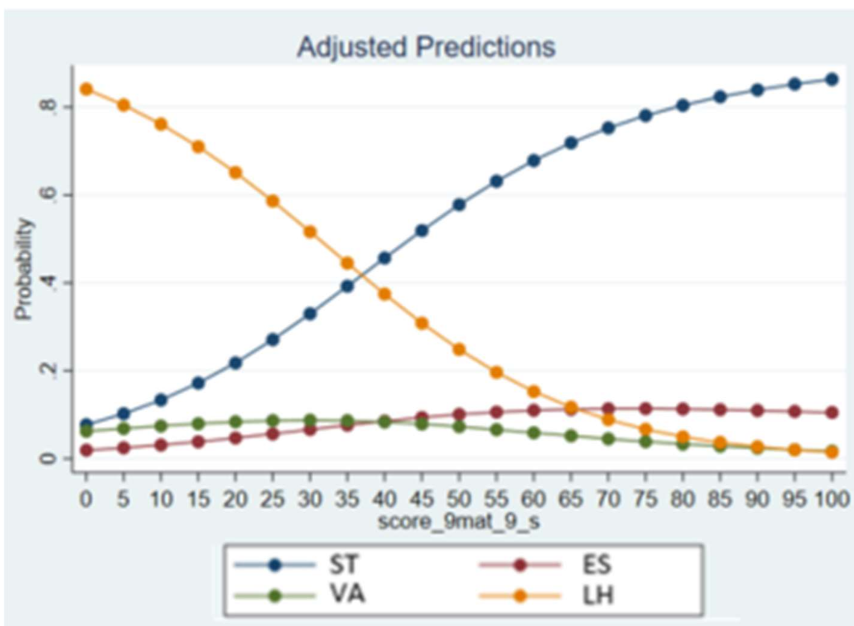
The SASE variable, that measures if students receive support from the State to study, has a positive coefficient for the VA and LH area and a negative coefficient for ST and ES area, meaning that having SASE support decreases the probability of following the areas with negative coefficient and contrarily increases the probability of following the areas with positive coefficient. The only variable that hadn't any statistically significant effect in the probability of following a certain area is the Portuguese variable.

The variables with more significant impact in our model to predict the probability to follow a certain area ultimately were the students' academic achievements with particular emphasis in the mathematics exam scores. An increase of 1 point, in a scale of 0 to 100, in the mathematics national exam leads to an increase of 0.81 p.p. of the probability to follow a ST area in detriment of the others. In the other direction the same increase results in the decrease of the probability to choose a LH area by 0.83 p.p..

These results are very significant and although we cannot imply direct causation using this model we can corroborate evidence of correlation between mathematic exam scores and the area chosen within the academic track. Students who have weak results in mathematics tend to prefer the areas with lower levels of mathematic relevance. The negative coefficient for the variable score_mat in the VA and particularly LH point to that. We can also look to these hypothesis in the inverse sense. The areas with higher required knowledge of Mathematics, which are ST and ES, contrasting with previous mentioned areas present, logically, a positive coefficient for the variable score_mat, suggesting that students with already higher achievements in this subject, through vocational appetence will prefer areas with high exigency of the mathematics subject.

In order to visualize this effect we have applied a marginal effects at the mean approach to see how students with all the explanatory variable set at mean values would choose the areas based on different results in the national mathematics exam (ceteris paribus context). *Figure 2*, below, reflects our results:

Figure 2. *Predicted probability at the mean of choosing each area, for model 2**,** given different Math exam scores (ceteris paribus)*



One can see that students with very low grades in mathematics have lower preference for areas mathematics oriented and as the exam grade increases, the probability of the student to choose an area more maths oriented, especially the ST one, increases substantially. Although this trend is not verified in the Portuguese exam, we can at some extent corroborate the findings of and

Schneider et al. (1998) regarding the apparent correlation between higher academic ability and the pursuit of STEM fields.

The Portuguese exam scores also present interesting results but not as significant as the mathematic exams. Each additional point in the Portuguese exams reflects an increase in the probability of following the LH area by 0.22 p.p. in average in relation to other areas. All the other areas, have negative coefficients meaning that a similar increase reduces the probability in average for these areas to be chosen.

6.3 Robustness Check

One of the main assumptions of a Multinomial Logistic regression is that there is independence of irrelevant alternatives (IIA). Hausman and McFadden (1984) suggested that if part of the choice set is irrelevant in respect to other alternatives, omitting that subset from the multinomial logit model will not lead to inconsistent estimates. Applying a Hausman or Small-Hsiao test allow to test for the validation of the IIA assumption, however the results represented a violation of this principle, see *Appendix 5 and 6*. There is a long-standing debate on the validity of these tests (e.g. Cheng and Long, 2006) but in order to confirm our estimates a Multinomial Probit regression was employed. The model allows to predict the likelihood of a student to choose between the four different areas while relaxing the IIA assumption. The AME were very similar and can be consulted in *Appendix 7*.

7. Discussion and Conclusions

When looking at the student's decision between an academic track and a professional track the more relevant factors of decision relate to the SES, academic ability and gender. These factors are more relevant for students with lower SES which may lead to further inequalities in the educational system. Students whose parents have lower education seem more sensitive to the effects of these variables, as model 1A and 1B showed, which implies that, for example, they will need to get significant better exam grades than students with higher parental education to have the same probability of following an academic track which in theoretical terms is more likely to bring success to their careers.

In relation to the decision within the academic track, while still relevant, variables that indicate the SES level seem to lose significance in relation to the prior analysis. The most predominant effect in the choosing of the area seems to be the academic ability with students appearing to select areas based on their level of mathematics score. This result suggests that the lack of mathematics attainment among students is the biggest obstacle for a career in a scientific field at secondary level. Gender also as significant impact in the probability of following a specific area showing evidence of gender-based patterns in decisions.

The lack of information regarding private schools may constitute a bias in our analysis, since students who pursue this sort of education represent in average students with significantly higher SES, therefore the sample misses a very specific group of our school population. In the first analysis the sample was divided in two groups, in order to analyse the different impact of the variables when considering students with more or less educated parents. With the difference of the coefficients among the groups being statistical significant, additional research could be conducted by dividing the sample in smaller groups and see if the difference in coefficients persisted. Also, in the second analysis a similar methodology could be employed. It also can be argued that using only 9th grade exam scores as proxy for academic ability is an insufficient measure, therefore in future analysis it could be employed classroom data to better evaluate academic ability. The main thing to note in the research is that regression results do not reflect direct causation. Therefore, as stated in the beginning of the project, these findings can be used only as a first step in trying to improve our educational system.

References

Ayalon, H. 2005. “Women and men go to university: Mathematical background and gender differences in choice of field in higher education”. Sex Roles: A Journal of Research, 48(5–6), 277–286.

Barker, L. J. and Garvin-Doxas, K. 2004. “Making Visible the Behaviors That Influence Learning Environment: A Qualitative Exploration of Computer Science Classrooms.” Computer Science Education, 14(1), 119–145.

Betts, J. R. 2011. “The Economics of Tracking in Education.” In Handbook of the Economics of Education, Volume, edited by E. A. Hanushek, S. Machin, and L. Woessmann,341–381. Vol. 3. Amsterdam: North Holland

Beyer, S., DeKeuster, M., Rynes, K. and DeHeer, J. 2004. “The temporal stability of gender differences in MIS students”. Proceedings of the 10th Americas Conference on Information Systems (pp. 1197–1204). New York, NY: AMCIS.

Blanden, J. 2005. “Love and Money: Intergenerational Mobility and Marital Matching on Parental Income”. Research Paper 272, Family and Labour Studies, Statistics Canada.

Blanden, J., Gregg, P. and Machin, S. 2005. “Intergenerational mobility in Europe and North America”. London: London School of Economics, Centre for Economic Performance.

Blanden, J., Gregg, P., and Macmillan, L. 2007. “Accounting for intergenerational Income Persistence: Non-cognitive Skills, Ability and Education”. Economic Journal, 117, C46-C60

Breen, R. and Goldthorpe, J. 1999. “Class inequality and meritocracy: a critique of Saunders and an alternative analysis”. British Journal of Sociology, 50:1, pp.1-27.

Breen, R. and Jonsson, J.O. 2005. “Inequality of opportunity in comparative perspective: recent research on educational attainment and social mobility”. Annual Review of Sociology, 31, pp.223-243.

Busch, T. 1995. “Gender differences in self-efficacy and academic performance among students of business administration”. Scandinavian Journal of Educational Research, 39, 311–318.

Butlin, George 1999. “Determinants Of Post-Secondary Participation.” Education Quarterly Review 5 (3): 9-35.

Coleman, J. S. 1987. “Families and schools”. Educational Researcher, 16(6), 32–38.

Cheng, S. and J. Scott L. J. 2006. “Testing for IIA in the Multinomial Logit Model.” Sociological Methods & Research: 35: 583-600.

Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions" Econometrica. 28 (3): 591–605

Dai, Y. 1996. “Educational Plans After High School: A National Survey Of High School Seniors." Journal Of Research And Development In Education 30 (1): 22-30.

Davis-Kean, P. E. 2005. “The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment”. Journal of Family Psychology, 19, 294–304.

DGEEC 2019. “Transição entre o ensino secundário e o ensino superior: 2016/17-2017/18”

DGEEC 2020. “Perfil do Aluno 2018/2019”.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C. and Mac Iver, D. 1993. “Development during adolescence: The impact of stage-environment fit on young adolescents - experiences in schools and in families”. American Psychologist, 48, 90–101.

Furr, L. A. 1998. "Fathers' Characteristics And Their Children's Scores On College Entrance Exams: A Comparison Of Intact And Divorced Families." Adolescence 33 (131): 533-42.

Gammie, E., Paver, B., Gammie, B. and Duncan, F. 2003. “Gender differences in accounting education: An undergraduate exploration”. Accounting Education, 12(2), 177–196.

Gilbert, S., Barr, L., Clark, W., Blue, M. and Sunter ,D. 1993. “Leaving School: Results From A National Survey Comparing School Leavers And High School Graduates 18 To 20 Years Of Age”. Human Resources Development Canada.

Grebennikov, L. and Skaines, I. 2009. "Gender and higher education experience: A case study". Higher Education Research & Development, 28(1): 71–84.

Hanushek, E. A. and Wössmann, L. 2006 “Does Educational Tracking Affect Performance and Inequality ? Differences-in-Differences Evidence across Countries.” Economic Journal, Royal Economic Society116 (510):C63–C76. 03.

Hausman, J. A. and McFadden, D. L. 1984. “Specification tests for the multinomial logit model”. Econometrica 52: 1219–1240.

Henriques, R. 2018. “Vocational education: coursetaking choice and impact on dropout and college enrollment rates" (Master’s thesis). Nova School of Business and Economics.

Heyns, B. 1974. "Social Selection and Stratification within Schools." American Journal of Sociology, 79: 1434-1451.

Hurtado, S., Inkelas, K. K., Briggs, C. and Rhee, B. S. 1997. "Differences in College Access and Choice among Racial/Ethnic Groups: Identifying Continuing Barriers." Research in Higher Education, 38 (1): 43-75.

Jencks, C., Smith, M., Ackland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B. and Mickelson, S. 1972. “Inequality: A Reassessment of the Effect of Family and School in America”. New York: Basic Books.

Kniveton, B. H. 2004. “The influences and motivations on which students base their choice of career”. Research in Education, 72,47-59

Kulik, J. 1998. "Curricular Tracks and High school Vocational Education". In The Quality of Vocational Education: Background Papers from the 1994 National Assessment of Vocational: 65-131.

Leme, A., Escardíbul, J. O., Nunes L. C., Reis, A.B., and Seabra, C. 2020. “The effect of a specialized versus a general upper-secondary school curriculum on students’ performance and inequality. A difference-in-differences cross-country analysis”. Applied Economics

Looker, D. E. 1997. "In Search of Credentials: Factors Affecting Young Adults’ Participation in Post-secondary Education." The Canadian Journal of Higher Education, 27 (1- 3).

Looker, D. E. 1993. “Interconnected Transitions And Their Costs: Gender And Urban / Rural Differences In The Transitions To Work.” In P. Anisef And P. Axlerod (Eds.) Schooling And Employment In Canada. Toronto.

Looker, D. E. 1994. “Active Capital: The Impact Of Parent’s On Youths’ Educational Performance And Plans.” In L. Erwin And D. Maclennan (Eds.), Sociology Of Education In Canada. Copp Clark Longman, P. 164-87.

Looker, D. E. and Lowe, G. 2001 “Post-Secondary Access and Student Financial Aid in Canada: Current Knowledge and Research Gaps”

Lowe, G., Krahn, H., and Bowlby,J. 1997. “1996 Alberta High School Graduate Survey: Report Of Research Findings”. Edmonton: Communications Branch, Alberta Advanced Education And Career Development, January.

Lubinski, D. and Benbow, C. P. 1992. “Gender differences in abilities and preferences among the gifted: Implications for the math/science pipeline”. Current Directions in Psychological Science, 1: 61-66.

Madigan, T. 1997. “Science proficiency and course-taking in high school: The relationship of science course-taking patterns to increases in science proficiency between 8th and 12th grades”. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Educational Statistics.

Maxwell, M. P. 1996. “Ethnicity, Gender, And Occupational Choice In Two Toronto Schools." Canadian Journal Of Education 21 (3): 257-79.

Mortensen, T. G. 1995. "The Mortensen Report on Public Policy Analysis of Opportunity for Postsecondary Education." Postsecondary Education Opportunity 31-42.

Neuman, S. and Ziderman, A. 1999. “Vocational education in Israel: Wage effects of the VocEd-Occupation match.” The Journal of Human Resources, 34(2): 407-420. ty of Vocational Education: Background Papers from the 1994 National Assessment of Vocational: 65-131.

Newfoundland 1998. “On The Road To Success: The Report Of The Follow-Up Of 1995 And 1996 High School Graduates”. Department Of Education.

Nunn, A., Johnson, S., Monro, S., Bickerstaffe, T. and Kelsey, S. 2007. “Factors influencin g social mobility”. Technical Report. Her Majesty’s Stationery Office.

Pearson, C. and Jenkins M. D. 1997. “Parental Influence On A Student’s Selection Of A College Major.” College Student Journal 71 (3): 301-13.

Perron, Jaques 1996. "Ethnicity and Educational Aspirations of High-School Students."In Galaway, Burt and Hudson, J. (eds.) Youth in Transition: Perspectives on Research and Policy. Thompson Educational Publishing, Toronto.

Ramos, Â. 2017. "The impact of choosing a vocational track on academic performance and college enrollment in Portugal" (Master’s thesis). Nova School of Business and Economics

Rasinski, K., and Pedlow, S. 1994. “Using Transcripts to Study the Effectiveness of Vocational Education.” Journal of Vocational Education Research, 19:23- 43.

Schneider, B., Swanson, C. B., & Riegle-Crumb, C. 1998. “Opportunities for learning: Course sequences and positional advantages”. Social Psychology of Education, 2, 25–53.

Seymour, E., and Hewitt, N. 1997. “Talking About Leaving: Why Undergraduates Leave the Sciences”. Boulder, Colo.: Westview Press, 1997.

Shepard, L., Schmidt, J. and Pugh, R. 1992. "Factors Influencing High School Students' Differences In Plans For Post Secondary Education: A Longitudinal Study." Paper Presented At The Annual Meeting Of The American Educational Research Association (San Francisco, CA, April 20-24).

Simpson, J. C. 2003. “Mom matters: Maternal influence on the choice of academic major”. Sex Roles: A Journal of Research, 48(9110),447-460.

Spence, M. 1973. "Job market signaling". Quarterly Journal of Economics, 87(3): 355- 74.

Strenta, A. C., Elliot, J., Adair, R., Matier, M., Scott, J. 1994 “Choosing and Leaving Science in Highly Selective Institutions.” Research in Higher Education, 1994, 35(5), 513–547.

Svoboda, R., Rozek, C., Hyde, Harackiewicz, J. and Destin, M. 2016 “Understanding the Relationship Between Parental Education and STEM Course Taking Through Identity-Based and Expectancy-Value Theories of Motivation”

Thiessen, V., and Looker, D. 1999. “Investing In Youth: The Nova Scotia School-To-Work Transition Project”. Human Resources Canada, Hull Quebec.

Trusty J., Robinson, C. R., Plata, M., and Ng, K.M. 2000. “Effects Of Gender, Socio-economic Status, And Early Academic Performance On Post-Secondary Education Choice.” Journal Of Counseling And Development 78 (4): 463-72.

Vetter, B. M. 1996. “Myths and Realities of Women’s Progress in the Sciences, Mathematics, and Engineering.” In C. Davis, A. B. Ginorio, C. S. Hollenshead, B. B. Lazarus, P. M. Rayman, and Associates (eds.), The Equity Equation: Fostering the Advancement of Women in the Sciences, Mathematics, and Engineering. San Francisco: Jossey-Bass.

Appendixes

Appendix 1- *Summary Statistics of Academic and Vocational track*

Variables	Academic Track	VET	|Differences in mean|
Number of obs.	327,265 (72.66%)	123,129 (27.34%)	-
Female (%)	54.66	39.87	14.79***
Portuguese (%)	93.59	91.62	1.97***
SASE (%)	32.46	51.21	18.75***
Internet (%)	71.88	62.33	9.55***
Computer(%)	78.97	73.33	5.64***
Educy_mom	10.62	7.84	2.78***
Educy_dad	9.73	7.16	2.57***
Educy_average	10.20	7.53	2.67***
Score_pt	58.00	45.03	12.97***
Score_mat	54.51	33.31	21.2***
Age	15.22	16.11	0.89***

t-test significance level
*** p<0.01, ** p<0.05,*p<0.1

Appendix 2- *Chow test between model 1A and model 1B*

$$\text{For model 1A } \mathrm{X\_1_i} = \begin{matrix} \mathrm{female_i} \\ \mathrm{Portuguese_i} \\ \mathrm{SASE_{it-1}} \\ \mathrm{Internet_{it-1}} \\ \mathrm{Computer_{it-1}} \\ \mathrm{Educy\_average_{it-1}} \\ \mathrm{Score\_pt_{it-1}} \\ \mathrm{Score\_mat_{it-1}} \\ \mathrm{Age_{it}} \\ \mathrm{Year1_{it}} \\ \cdots \\ \mathrm{Year5_{it}} \\ \mathrm{District1_{it}} \\ \cdots \\ \mathrm{District17_{it}} \end{matrix} \quad \text{and for model 1B } \mathrm{X\_2_i} = \begin{matrix} \mathrm{female_i} \\ \mathrm{Portuguese_i} \\ \mathrm{SASE_{it-1}} \\ \mathrm{Internet_{it-1}} \\ \mathrm{Computer_{it-1}} \\ \mathrm{Educy\_average_{it-1}} \\ \mathrm{Score\_pt_{it-1}} \\ \mathrm{Score\_mat_{it-1}} \\ \mathrm{Age_{it}} \\ \mathrm{Year1_{it}} \\ \cdots \\ \mathrm{Year5_{it}} \\ \mathrm{District1_{it}} \\ \cdots \\ \mathrm{District17_{it}} \end{matrix}$$

With algebraic modifications rewrite the model as:

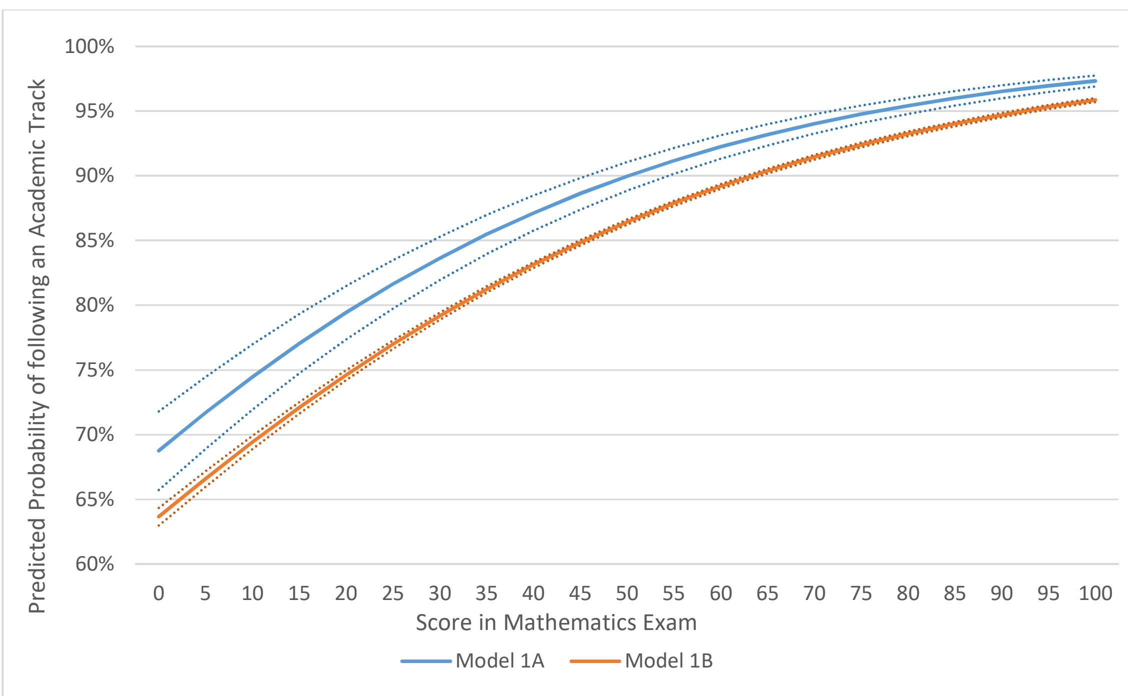
$$\mathrm{Academic_{it}} = \mathrm{X_i} * \beta_1 + \mathrm{d_2} * X_{2_i} * (\beta_2 - \beta_1) + \mathrm{d_1} * \varepsilon_{1i} + \mathrm{d_2} * \varepsilon_{2i}$$

Where X = (X_1, X_2)

Now it is possible to test if the coefficients are equal across the two models

```
qui test _b[Female1_2]=0, notest

qui test _b[PT_2]=0, accum notest

qui test _b[sase1_9_s_2]=0, accum notest

qui test _b[internet2_9_s_2]=0, accum notest

qui test _b[computer2_9_s_2]=0, accum notest

qui test _b[educy_average_2]=0, accum notest

qui test _b[score_9pt_9_s_2]=0, accum notest

qui test _b[score_9mat_9_s_2]=0, accum notest

qui test _b[age_2]=0, accum notest

test _b[group2]=0, accum

( 1)  [regular]Female1_2 = 0
( 2)  [regular]PT_2 = 0
( 3)  [regular]sase1_9_s_2 = 0
( 4)  [regular]internet2_9_s_2 = 0
( 5)  [regular]computer2_9_s_2 = 0
( 6)  [regular]educy_average_2 = 0
( 7)  [regular]score_9pt_9_s_2 = 0
( 8)  [regular]score_9mat_9_s_2 = 0
( 9)  [regular]age_2 = 0
(10)  [regular]group2 = 0

           chi2( 10) =  178.56
         Prob > chi2 =    0.0000
```

Appendix 3 - *Predicted probabilities at the mean of choosing the academic track for model 1A and 1B given different Math exam scores (ceteris paribus)*



__Appendix 4-__ Summary Statistics of the areas within the Academic Track

Variables	ST	ES	VA	LH
Number of obs.	182,688 (55.82%)	34,288 (10.48%)	23,554 (7.20%)	86,735 (26.50%)
Gender (%)	49.82	46.81	65.07	65.15
Portuguese (%)	94.54	93.36	91.02	92.37
SASE (%)	28.57	26.83	36.73	41.66
Internet (%)	73.82	71.12	69.89	68.62
Computer(%)	80.69	76.81	77.17	76.63
Educy_mom	11.11	11.27	10.26	9.41
Educy_dad	10.16	10.41	9.42	8.61
Educy_average	10.66	10.86	9.87	9.04
Score_pt	60.70	57.14	53.55	53.69
Score_mat	63.24	58.30	45.29	36.56
Age	15.12	15.21	15.43	15.38

__Appendix 5 -__ Testing the IIA Assumption using Hausman Tests

Hausman tests of IIA assumption (N=256383)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	chi2	df	P>chi2
ST	837.396	62	0.000
ES	1877.752	62	0.000
VA	206.524	62	0.000
LH	190.714	63	0.000

Note: A significant test is evidence against Ho.

suest-based Hausman tests of IIA assumption (N=256383)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	chi2	df	P>chi2
ST	574.429	64	0.000
ES	603.420	64	0.000
VA	332.731	64	0.000
LH	531.960	64	0.000

Note: A significant test is evidence against Ho.

__Appendix 6 -__ Testing the IIA Assumption using a Small-Hsiao Test

Small-Hsiao tests of IIA assumption (N=256383)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	lnL(full)	lnL(omit)	chi2	df	
ST	-4.38e+04	-4.38e+04	115.051	64	0.000
ES	-7.69e+04	-7.68e+04	88.399	64	0.023
VA	-8.64e+04	-8.64e+04	88.855	64	0.022
LH	-5.93e+04	-5.93e+04	70.269	64	0.276

Note: A significant test is evidence against Ho.

***Appendix** 7- Results for model 2 using Multinomial Probit: AME predicting participation in the different areas of the academic track.*

	2			
	ST	ES	VA	LH
Gender	-0.0991*** (0.0018)	-0.0216*** (0.0012)	0.0311*** (0.0010)	0.0896*** (0.0015)
Portuguese	-0.0044 (0 .0044)	0.0018 (0.0030)	-0.0017 (0.0022)	0.0043 (0.0036)
SASE	-0.0094*** (0.0020)	-0.0096*** (0.0014)	0.0045*** (0.0011)	0.0145*** (0.0017)
Internet	-0.0057* (0.0033)	0.0014 (0.0023)	0.0034* (0.0019)	0.0009 (0.0028)
Computer	0.0115*** (0.0038)	-0.0081*** (0.0026)	-0.0050** (0.0021)	0.0015 (0.0032)
Educy_average	-0.0012*** (0.0003)	0.0018*** (0.0002)	0.0020*** (0.0001)	-0.0026*** (0.0002)
Score_pt	-0.0004*** (0.0001)	-0.0014*** (0.0000)	-0.0006*** (0.0000)	0.0023*** (0.0001)
Score_mat	0.0082*** (0.0000)	0.0009*** (0.0000)	-0.0007*** (0.0000)	-0.0083*** (0.0000)
Age	-0.0723*** (0.0019)	0.0112*** (0.0013)	0.0275*** (0.0009)	0.0336*** (0.0015)
Number of obs.	256,383			